Quantitative Modeling of Soil Sorption for Xenobiotic Chemicals

by Aleksandar Sabljić*

Experimentally determining soil sorption behavior of xenobiotic chemicals during the last 10 years has been costly, time-consuming, and very tedious. Since an estimated 100,000 chemicals are currently in common use and new chemicals are registered at a rate of 1000 per year, it is obvious that our human and material resources are insufficient to experimentally obtain their soil sorption data. Much work is being done to find alternative methods that will enable us to accurately and rapidly estimate the soil sorption coefficients of pesticides and other classes of organic pollutants.

Empirical models, based on water solubility and n-octanol/water partition coefficients, have been proposed as alternative, accurate methods to estimate soil sorption coefficients. An analysis of the models has shown (a) low precision of water solubility and n-octanol/water partition data, (b) varieties of quantitative models describing the relationship between the soil sorption and above-mentioned properties, and (c) violations of some basic statistical laws when these quantitative models were developed.

During the last 5 years considerable efforts were made to develop nonempirical models that are free of errors imminent to all models based on empirical variables. Thus far molecular topology has been shown to be the most successful structural property for describing and predicting soil sorption coefficients. The first-order molecular connectivity index was demonstrated to correlate extremely well with the soil sorption coefficients of polycyclic aromatic hydrocarbons (PAHs), alkylbenzenes, chlorobenzenes, chlorinated alkanes and alkenes, heterocyclic and heterosubstituted PAHs, and halogenated phenols. The average difference between predicted and observed soil sorption coefficients is only 0.2 on the logarithmic scale (corresponding to a factor of 1.5). A comparison of the molecular connectivity model with the empirical models described earlier shows that the former is superior in accuracy, performance, and range of applicability.

It is possible to extend this model, with the addition of a single, semiempirical variable, to take care of polar and ionic compounds and to accurately predict the soil sorption coefficients for almost 95% of all organic chemicals whose coefficients have been reported. No empirical or nonempirical models have ever predicted the soil sorption coefficients to such a high degree of accuracy on such a broad selection of structurally diverse compounds. An additional advantage of the molecular connectivity model is that it is sufficient to know the structural formulas to make predictions about soil sorption coefficients. The structural analysis of our quantitative model has shown that two factors are responsible for the majority of the variations in the soil sorption data: the molecular surface areas and the polarity of compounds.

Distribution of Xenobiotic Chemicals in the Environment

The widespread use of organic pesticides has given rise to extensive interest (1–3) in the adsorption of such solutes by soils from aqueous solutions because of the influence of this process on pesticide performance, mobility in the soil, and residue problems. Contamination of groundwater by pesticides and other agricultural chemicals by hazardous chemicals from waste disposal sites, and by gasoline and chemicals from underground storage tanks is becoming a major environmental problem. Although universal, this problem is particularly emphasized in the United States and other industrialized countries as well. Reports prepared by the U.S.

Environmental Protection Agency (4) and the Council on Environmental Quality (5) indicate that up to 50,000 waste disposal sites in the United States may contain hazardous chemicals. Groundwater systems close to many of these sites are being slowly degraded, and the contamination often involves the presence of synthetic organic materials (6).

To minimize the impact of man's activity on groundwater quality, mechanisms by which pollutants enter groundwater need to be better understood, and reliable techniques either to measure or predict the transport of contaminants within aquifers need to be developed. Because the large majority of synthetic organic chemicals are hydrophobic, their adsorption on soil, sediments, or other materials plays a very important role in their transport in surface or subsurface systems.

Present laws and regulations about control of toxic substances (e.g., Toxic Substances Control Act, U.S. Federal Pesticides Law, OECD-Guideline for Testing

*Theoretical Chemistry Group, Department of Physical Chemistry, Institute Rudjer Bošković, HPOB 1016, 41001 Zagreb, Croatia, Yugoslavia.

Chemicals) require that all compounds that are prepared commercially or identified as potentially useful chemicals must be assessed for their environmental behavior and hazards. Many compounds (~ 100,000) (7) are currently in common use, and new chemicals are registered at a rate of 1000 per year. Our human and material resources are not sufficient to obtain experimentally even basic information about the environmental fate for all these compounds. Thus, it is necessary to develop quantitative models that will reliably and rapidly predict the environmental behavior for such large sets of compounds.

The environmental fate of organic pollutants depends strongly on their distribution between different environmental compartments. The soil sorption coefficients are currently used as a quantitative measure of the adsorption of xenobiotic chemicals by soil from aqueous solutions (2). They are defined as the ratios between the concentrations of a given chemical sorbed by the soil and dissolved in soil water. In order to compare the soil sorption coefficients measured for different soils, they have to be normalized either to the total organic carbon content of the soil (K_{oc}) or the organic matter content of the soil (K_{om}). These two normalizing schemes are simply related by the factor 1.724; thus, it is easy to convert coefficients reported on any basis.

In this paper, quantitative models for predicting soil sorption coefficients of xenobiotic chemicals will be described and evaluated. A small section will be devoted to general principles of statistical modeling and to QSAR models in particular. Another purpose of this paper is to shed more light on the mechanism(s) of the soil sorption process since it is currently described either as an adsorption or a partitioning process. The details about this controversy will be discussed in the section discussing the mechanism of the soil sorption process. In the closing section we shall describe the applications for quantitative models of soil sorption and give our views about needs and future developments in this environmental science field.

Principles and Rules of Statistical Modeling

General Considerations on Statistical Modeling

Before proceeding any further, we shall discuss some elementary principles and pitfalls of modeling processes in general and statistical modeling in particular. While the following points should be self-evident, they have generally been ignored and/or overlooked. On the whole, models are used in the sciences to represent, explain, predict, or estimate phenomena of interest. They are simply approximations of real systems primarily developed for the purpose of prediction, and we cannot expect them to be the true descriptions of reality. Thus, the real value of a model is related neither to the type and size of the model nor to its results for the training data set, but rather to its ability to handle new situations correctly. Yet, as Ptolemy demonstrated, bad models are as readily available as the good ones. Moreover, microcomputers have made this era the golden age of modeling, and easy access to desktop computers and modern software allows people with either modest or no statistical background to construct and test elaborate formal models. In addition, microcomputers enable us to abuse models at superhuman speed and to produce enormous volumes of questionable numerical results. Thus, the principles of logic, statistics, and measurement must be satisfied in order to develop a meaningful model and to successfully predict phenomena of interest. The data sets must conform with the basic statistical requirements underlying the statistical procedure used to develop a particular model. Also, the model must make logical, mechanistic, and statistical sense. Particular caution should be exercised not to overfit the data. As demonstrated later, many quantitative environmental models are published that do not comply with the above basic requirements.

Statistical Modeling Process

There are five major stages in the process of statistical modeling: identification, fitting and estimation, validation, application, and iteration. Figure 1 gives the schematic description of this process.

Identification is the process of finding or choosing an appropriate model for a particular situation. There are no rigorous procedures that guarantee success. Systematically, there are two extreme approaches in the identification process: one which seeks a model on the basis
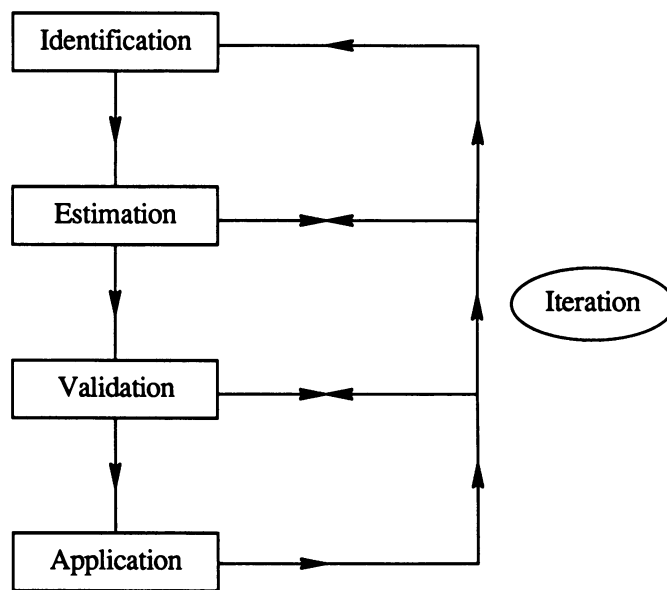

FIGURE 1. The process of statistical modeling.

of a rational argument from some knowledge of the real life situation without reference to any actual data (conceptual identification), and the other that considers only the data and their properties (empirical identification). In practice, both approaches should be used and combined to create the best models possible.

Model fitting is the stage when we move from the general form to the specific numerical form, whereas the estimation stage represents the process of assigning numerical values to parameters in that model. The most frequently used and the best known method for the purpose of statistical modeling is least squares fit.

The process of comparing the model with the observed world is called "validation." What is valid at one stage of a study need not be valid at a more developed stage. A model can be valid for one purpose, but not for another. The object of validation is to examine whether the model is a good (not true) description of reality in terms of its behavior and of its intended application(s).

For logical reasons and convenience of presentation, the application is described as the last stage of the modeling process. However, in practice, the application must be a part of the very first consideration and taken into account when carrying out the processes of identification, estimation, and validation, since models are developed to help solve problems. If we ignore application during the modeling process, we might end up with an excellent model that will not solve our problem.

Figure 1 gives a realistic view of statistical modeling being an iterative process. It is a process of continuous development, going back a stage or two to use additional information. The model is never "the model," final and unaltered. It is always a tentative model, which we shall use until we can improve it.

We should always remember that in statistical modeling we are dealing with probabilities, distributions, populations, and uncertainties. One of the major sources of uncertainty is our data. When seeking to fit models to our data, we often find that an accurate fit requires more data than we can either practically or financially obtain. Even when we have enough data, we find that its quality is often far from perfect. We may have accurate data on some variables, whereas those for other variables may be inadequate. Very often, in order to have enough data, we are forced to collect data from various sources. The reliability and quality of such data will vary greatly. The limitations on the amount and quality of the data available reduce the precision with which we can fit and use the models.

Quantitative Structure-Activity Relationships (QSAR) Models

Mathematical models that relate some chemical, biological, or environmental activity of interest to some quantitative structural descriptor or physico-chemical property are collectively known as quantitative structure-activity relationships (QSAR) models. QSAR models are usually developed for a group(s) of structural congeners. The primary objective in creating them is to predict the activities of untested congeners. The investigators also hope to understand better the mechanisms of action of structures under study.

The statistical procedure used to derive QSAR models is the linear regression analysis, and it can be either single or multivariable, depending on the number of structural descriptors used in the particular analysis. The usual procedure in deriving QSAR models is stepwise and begins with the single variable regressions going from the simplest to the more complex structural descriptors. The next step is to screen multivariable models of increased complexity until the simplest model predicting activity of interest within the experimental error is found. Naturally, this stepwise procedure can be discontinued with the single variable models. To test the quality and accuracy of derived models, the following statistical parameters should be used: the single (r) and multiple (R) correlation coefficients, the standard error of the estimate (s), a test of null-hypothesis (F-test), and the amount of explained variance (EV).

Several major mistakes are commonly made in today's environmental QSAR modeling: a) the principle of simplicity is ignored; b) experimental error in the dependent variable is neglected and models are overfitted, c) too many independent variables are used with small sets of compounds and with a high probability of chance correlation; d) models are often made for a small congeneric series, checked and validated only with basic statistical tests (correlation coefficient and standard deviation), and residuals are rarely analyzed for possible patterns, and e) models are rarely validated and never against the entire knowledge base and data base.

Modeling the Soil Sorption Coefficients: Past Results with Empirical Variables, Their Limitations and Present Needs

Considerable work has already been done to find alternative methods that will enable us to accurately and rapidly estimate and/or predict soil sorption coefficients of xenobiotic chemicals (2, 8–13). Water solubility (WS) and n-octanol/water partition coefficient (K_{ow}) are generally accepted as accurate estimates of the soil sorption coefficients. The quantitative models describing relationships between the soil sorption coefficients and WS or K_{ow} are shown in Tables 1 and 2, respectively.

Our detailed analysis of these models (14) showed surprisingly high variability in reported WS and K_{ow} data and quantitative linear models describing their relationship to the soil sorption coefficients. For the sake of clarity, the results and models with each of the above empirical variables will be discussed separately.

The low precision of the K_{ow} data is illustrated in Table 3.

A brief examination of Table 3 clearly shows a surprisingly high variability in experimental K_{ow} data,

Table 1. Published quantitative models describing correlations between observed soil sorption coefficients and n-octanol/water partition coefficients.

	Equation	Reference
a)	$\log K_{om} = 1.00 \log K_{ow} - 0.45$ $N = 10 \quad r = 1.00$	(12)
b)	$\log K_{om} = 1.00 \log K_{ow} - 0.56$ $N = 22 \quad r^2 = 0.98$	(13)
c)	$\log K_{om} = 0.72 \log K_{ow} + 0.25$ $N = 13 \quad r^2 = 0.95$	(10)
d)	$\log K_{om} = 0.82 \log K_{ow} - 0.22$ $N = 9 \quad r^2 = 0.98$	(11)
e)	$\log K_{om} = 0.90 \log K_{ow} - 0.78$ $N = 12 \quad r^2 = 0.989$	(3)
f)	$\log K_{om} = 0.54 \log K_{ow} + 1.14$ $N = 45 \quad r^2 = 0.74$	(2)
g)	$\log K_{om} = 0.52 \log K_{ow} + 0.64$ $N = 105 \quad r^2 = 0.90$	(9)

Table 2. Published quantitative models describing correlations between observed soil sorption coefficients and water solubility (in μmole/L).

	Equation	Reference
a)	$\log K_{om} = -0.55 \log WS + 3.40$ $N = 106 \quad r^2 = 0.71$	(2)
b)	$\log K_{om} = -0.54 \log WS + 3.44$ $N = 10 \quad r^2 = 0.94$	(12)
c)	$\log K_{om} = -0.56 \log WS + 4.04$ $N = 15 \quad r^2 = 0.988$	(8)
d)	$\log K_{om} = -0.58 \log WS + 4.00$ $N = 15 \quad r^2 = 0.635$	(17)
e)	$\log K_{om} = -0.81 \log WS + 3.87$ $N = 12 \quad r^2 = 0.995$	(3)
f)	$\log K_{om} = -0.69 \log WS^a + 4.27$ $N = 22 \quad r^2 = 0.933$	(13)

[a]Water solubility units are mg/L.

Table 3. Range of reported experimental n-octanol/water partition coefficients (log K_{ow}) for some alkylbenzenes, chlorophenols, PAHs, chlorobenzenes, PCBs, and other chlorinated hydrocarbons.[a]

Compound	Range of log K_{ow}
Toluene	2.11–2.73
1-Naphthol	2.31–2.98
2,4-Dichlorophenol	2.80–3.23
2,4,6-Trichlorophenol	3.62–4.05
2,4,5-Trichlorophenol	3.70–4.19
3,4,5-Trichlorophenol	3.69–4.41
2,3,4,5-Tetrachlorophenol	4.42–4.87
Pentachlorophenol	3.32–5.86
Trichloroethene	2.29–3.30
Benzene	1.56–2.15
Naphthalene	3.01–4.70
Chlorobenzene	2.18–3.79
1,3-Dichlorobenzene	3.24–3.60
1,2,4-Trichlorobenzene	3.93–4.67
1,2,3,5-Tetrachlorobenzene	4.46–4.94
Pentachlorobenzene	4.88–5.69
Hexachlorobenzene	4.13–7.42
2-Chlorobiphenyl	3.90–4.59
2,2'-Dichlorobiphenyl	4.04–5.00
2,4,5,2',4',5'-Hexachlorobiphenyl	6.34–8.18
p,p'-DDT	3.98–6.36
1,2-Dichloroethane	1.45–1.79
1,1,2,2-Tetrachloroethane	2.39–3.01
Tetrachloroethene	2.53–2.88
Aldrin	5.52–7.40

[a]Original literature is cited by Sabljić (14).

even for highly soluble compounds in water. The ranges of reported experimental K_{ow} data vary between 0.5 and 3.3 on the logarithmic scale. The accuracy of predicted soil sorption coefficients cannot be better than the accuracy of the experimental K_{ow} data. Table 3 contains only experimental K_{ow} data, and all reports cited have been published within the last several years.

The second difficulty in using the n-octanol/water partition coefficient to predict the soil sorption coefficients is the wide variety of reported quantitative linear models describing their relationship (2,3,8–13). This large diversity in published models is illustrated in Table 1. The slopes of the models range from 0.52 to 1.00 and their intercepts, from −0.78 to 1.14. The large diversity in log K_{om} versus log K_{ow} models has been recognized by other investigators in the field (15–17), but only one study (17) gave this problem serious and detailed consideration.

Notwithstanding these problems, the soil sorption coefficients were calculated for a test sample of 31 compounds (alkyl- and chloro-benzenes, heterocyclic and substituted polycyclic aromatic hydrocarbons, chlorinated alkanes and alkenes, and chlorinated phenols) with the quantitative models a, b, e, f, and g shown in Table 1 (14). (The identical test sample will be used to

test all models discussed in this review.) These five quantitative models were selected because they cover at least three orders of magnitude in the soil sorption data. When several K_{ow} coefficients were reported for the same compound, they were all used in the calculation. The predicted value for the soil sorption coefficient of the particular compound was expressed as the range of calculated values. The average range of predicted soil sorption coefficients from their K_{ow} coefficients was over 1.5 on the logarithmic scale (corresponding to a factor of 35). This result is consistent with the low accuracy of the experimental K_{ow} data.

Such variability in the predicted soil sorption coefficients shows that quantitative empirical models are highly dependent on the empirical data used to create them. Consequently, the soil sorption coefficients calculated by such models will depend on the empirical input data and the particular quantitative model used to calculate them. Currently, the empirical models based on n-octanol/water partition coefficients appear insufficiently reliable in predicting the environmental distribution of chemicals.

In addition to the coefficients' low accuracy, some basic statistical requirements were overlooked when quantitative models based on K_{ow} coefficients were developed. All quantitative models discussed above were derived through the statistical procedure known as the linear regression model or the method of least squares. The most important assumption underlying the linear regression model is that the dependent variable (commonly denoted as Y) contains all the errors in each data

pair (X,Y). In other words, the independent variable (commonly denoted as X) is free of error and has zero variance *(18,19)*. It is obvious from an examination of Table 3 that the data for n-octanol/water partition coefficients do not conform with this basic assumption. It was demonstrated some time ago (20) that if this basic assumption is violated, the fitted slopes can deviate by as much as 40% from the correct value. Thus, the validity and applicability of published quantitative models describing the relationship between the K_{ow} coefficients and soil sorption coefficients are highly questionable.

The n-octanol/water partition coefficient data calculated by the Hansch *(21)* or Rekker *(22)* methods will have the same level of accuracy as the experimental K_{ow} coefficients. They will contain almost all their experimental errors and uncertainties because both these methods are based on measured K_{ow} data. For example, the use of these methods for polycyclic aromatic hydrocarbons with four or more rings results in their calculated K_{ow}'s *(16,23)* being up to one order of magnitude higher than those measured by Means et al. *(24)*.

All that has been said for empirical models based on the n-octanol/water partition coefficients is also true for models based on water solubilities: a) the low precision of the measured water solubility data (Table 4), b) the wide variety of quantitative linear models correlating soil sorption coefficients and water solubility (Table 2) *(3,8,9,12,13,25)*, and c) the violation of statistical requirements for linear regression models by large errors in the experimental water solubility data. Because some quantitative models *(13,25,26)* use nonmolar units (mg/mL) for water solubility data, their use is further impeded.

From the discussion in this section, it is obvious that an accurate, simple, and fast model for predicting soil sorption coefficients of xenobiotic chemicals is still badly needed. During the last several years, considerable work has been done to find alternative, structurally

Table 4. Range of reported experimental water solubilities (mole/L) for some chlorobenzenes, alkylbenzenes, PAHs, PCBs, and other chlorinated hydrocarbons.[a]

Compound	−log WS range
1,2,3-Trichlorobenzene	3.76 − 4.17
1,2,3,4-Tetrachlorobenzene	4.25 − 5.31
1,2,4,5-Tetrachlorobenzene	4.56 − 5.86
Toluene	1.77 − 2.25
n-Butylbenzene	3.62 − 4.06
Dibenzothiophene	5.10 − 5.54
Benzene	1.64 − 1.98
Pyrene	6.18 − 6.80
1,4-Dichlorobenzene	3.21 − 3.68
1,2,4-Trichlorobenzene	3.57 − 5.78
Pentachlorobenzene	5.27 − 6.27
Hexachlorobenzene	6.78 − 7.78
2,5,2′,5′-Tetrachlorobiphenyl	6.80 − 7.69
2,4,5,2′,4′,5′-Hexachlorobiphenyl	8.47 −11.44
2,3,4,2′,3′,4′-Hexachlorobiphenyl	7.56 − 9.11
p,p′-DDT	7.95 −11.65
Tetrachloroethene	2.53 − 2.92
Lindane	4.50 − 6.29

[a]Original literature is cited by Sabljić *(14)*.

based methods that will enable us to accurately and rapidly estimate and/or predict the soil sorption coefficients of xenobiotic chemicals *(14,27-29)*. Detailed discussion about these methods and created quantitative models is given in the following sections.

Molecular Connectivity Models for Predicting Soil Sorption Coefficients

Our long-term interest *(14,27-29)* in environmental QSAR modeling is in developing general nonempirical model(s), based on information from structural formulas only, for predicting soil sorption coefficients of xenobiotic chemicals. The ultimate goal is to develop the above quantitative models with the following desired characteristics: *a)* be highly accurate for reliable use; *b)* be sufficiently simple, thus easily applicable by various scientists and other people (even laymen) involved in environmental problems as daily routine, either in the laboratory or in the field; and *c)* be short enough to be performed on large samples in a reasonable amount of time.

The nonempirical approach was selected because it has two major advantages over empirical approaches. First, the determination of nonempirical variables is faster and less expensive than the measurement of empirical variables and can be performed almost anywhere (e.g., office, laboratory, home, field, etc.), whereas measurements commonly require laboratory facilities with specialized equipment and qualified, experienced personnel. The second advantage is the higher accuracy of nonempirical variables which, by definition, have zero intrinsic error. This makes them ideally suitable within the rules of the standard linear regression model.

On the other hand, empirical variables (i.e., water solubility and the n-octanol/water partition coefficient) contain large experimental errors and uncertainties. Thus, it is necessary to use complex statistical procedures and experienced people to create a valid and meaningful quantitative relationship between experimental variables and environmental characteristics of hazardous chemicals.

The simplest way to represent a molecular structure is to assign it a number or a set of numbers, termed "indices." Indices generated by applying the chemical graph theory *(30)* are called "topological indices" *(30,31)*. The topological indices numerically express the topology of a chemical species and usually reflect, in varying degrees, their shape and size *(30,31)*. The particular advantage of topological indices is that they are direct and simple numerical descriptors of molecular structures and are used in the quantitative correlations with physical, chemical, biological, or environmental properties of molecules. The majority of topological indices are related either to adjacency relationship (atom-to-atom connectivity) or to topological distance (the number of bonds between the pair of atoms) in the molecular structure (chemical graph). Thus, they can be calculated either from the adjacency or from the distance matrix

of a chemical graph; the means by which this is accomplished varies from index to index.

The concept of molecular connectivity indices was introduced by Randić (32) and further developed and extensively exploited by Kier and Hall (33) and many others (14,27–31,34–51). Several extensive reviews of the theory and method of calculation of molecular connectivity indices have been published recently (30,31,33,52,53). Thus, only a brief description of the calculation of the first-order molecular connectivity index used in the nonempirical models in this study is given here.

The first-order ($^1\chi$) molecular connectivity indices are calculated from the nonhydrogen part of the molecule. Each nonhydrogen atom is described by its atomic δ value, which is equal to the number of adjacent nonhydrogen atoms. The $^1\chi$ are then calculated from the atomic δ values by Eq. (1):

$$^1\chi = \sum (\delta_i \, \delta_j)^{-0.5} \tag{1}$$

where i and j correspond to the pairs of adjacent nonhydrogen atoms and the summation runs over all bonds between nonhydrogen atoms.

Molecular connectivity indices for models discussed in this section were calculated by the GRAPH III computer program on an IBM PC/XT personal computer (31). Minimum hardware and software requirements for this program are an IBM PC or compatible computer, 256 KB of memory, 1 double-sided/double-density disk drive, and a PC-DOS or MS-DOS operating system, version 2.1 or higher. The use of a mathematical coprocessor is highly recommended. Its present version, GRAPH III, can calculate the molecular connectivity indices up to the tenth order for molecules with 35 nonhydrogen atoms or less. The program can be extended to handle larger molecules if sufficient memory is available.

During the last 5 years, much work has been done to develop nonempirical models that will be free from errors imminent to all models based on empirical variables. Thus far, molecular topology has been shown (14,27–29,40,41) to be the most successful structural variable for describing and predicting the soil sorption coefficients.

Several years ago we pioneered the idea that molecular topology could be successfully applied in correlations between molecular structure and the environmental distribution of xenobiotic chemicals (27,34). Subsequently, molecular connectivity indices were found to be very important structural variables for describing and predicting the soil sorption coefficients (27). In this preliminary study our attention was focused on polycyclic aromatic hydrocarbons (PAHs). They constitute the major group of environmental hazards to all living species because of their a) large production and widespread use, b) resistance to chemical and biological degradation, c) ability to accumulate in food chains and affect the growth and reproduction of organisms at all

levels of the food chain, and d) ability to cross the blood/brain barrier. Moreover, the soil sorption data of PAHs are from a single source (12), measured under highly controlled and uniform conditions, and therefore they have the highest degree of internal consistency and comparability. Thus, the single-source data was an excellent starting point for our project, since it gave us optimal control over the modeling process, good feeling for modeling environmental properties, and an excellent chance to create a sound model for predicting soil sorption coefficients.

The first-order molecular connectivity index was found to correlate extremely well with the soil sorption coefficients of eight PAHs as shown by Eq. (2) and its statistics.

$$\log K_{om} = 0.63 \, ^1\chi - 0.10$$

$$n = 8 \qquad r = 0.986 \qquad s = 0.202 \tag{2}$$
$$F^{1,6} = 205 \qquad \qquad EV = 96.7\%$$

This result was confirmed by two other laboratories (40,41). Such gratifying results were a driving force to our continuing investigation in the same direction.

The soil sorption coefficients were collected for 29 additional compounds, mainly halogenated hydrocarbons: chlorobenzenes, polychlorinated biphenyls, and chlorinated or brominated alkanes and alkenes. Adding halogen atoms to the hydrocarbon skeleton seemed to be a very small perturbation, and it is reasonable to expect that resulting compounds will have similar environmental distribution patterns as parent compounds. We were fortunate to learn that our working hypothesis is correct. A quantitative model describing the soil sorption of hydrocarbons and their halogenated derivatives is given by Eq. (3) and its statistics.

$$\log K_{om} = 0.53 \, ^1\chi + 0.42$$

$$n = 37 \qquad r = 0.976 \qquad s = 0.300 \tag{3}$$
$$F^{1,35} = 704 \qquad \qquad EV = 95.1\%$$

Statistically, Eq. (3) results account for 95% of the variation in the log K_{om} data. This variation is as good as can be expected, taking into account the accuracy of measured data. The alternative models log K_{om} versus log WS and log K_{om} versus log K_{ow} were also examined, and both were found to be inferior to Eq. (3). The fit for these empirical variables was 85 and 77% of the variation in the log K_{om} data, respectively.

Our next task was to expand our molecular connectivity model [Eq. (3)], to define the whole range of its applicability, and test the level of accuracy for predicting soil sorption coefficients. For this test we used the identical set of 31 compounds (alkyl- and chloro-benzenes, heterocyclic and substituted polycyclic aromatic hydrocarbons, chlorinated alkanes and alkenes, and chlorinated phenols) as previously used for testing models based on empirical variables (the previous section). Their first-order molecular connectivity indices

were calculated, and their soil sorption coefficients were predicted from the molecular connectivity model, Eq. (3).

A comparison of the observed and predicted sorption coefficients clearly demonstrates that the molecular connectivity model is very accurate in predicting the soil sorption coefficients. The average difference between predicted and observed soil sorption coefficients is only 0.22 on the logarithmic scale (corresponding to a factor 1.66), and more than 90% of the coefficients are predicted within the two standard deviations. (Only the 1,2,3,4- and 1,2,4,5-tetrachlorobenzenes and 2,3,4,5-tetrachlorophenol soil sorption coefficients are predicted outside the two standard deviation range.) This result is far superior to the empirical models, where the average difference between predicted and observed soil sorption coefficients is over 1.5 on the logarithmic scale (corresponding to a factor of 35).

The highly satisfactory performance of the molecular connectivity model Eq. (3) in predicting the soil sorption coefficients prompted us to combine all compounds, except the three outliers indicated above, into a single regression model. The resulting molecular connectivity model for the quantitative description of soil sorption coefficients of nonionic organic compounds is given in Eq. (4):

$$\log K_{om} = 0.53\ ^1\chi + 0.43$$

$$\begin{aligned} &n = 72 &&r = 0.977 &&s = 0.282 \\ &F^{1,70} = 1478 &&EV = 95.4\% \end{aligned} \tag{4}$$

The statistical parameters show that Eqs. (3) and (4) are statistically significant above the 99% level and both have similar levels of accuracy. Thus, the range of applicability of the molecular connectivity model is now extended to alkylbenzenes, heterocyclic PAHs, substituted PAHs, and chlorophenols without sacrificing high accuracy. In addition to extending the range of applicability of the molecular connectivity method, the above results show that the initial set of experimental data (*28*) is balanced: compare the regression coefficients and statistical parameters of Eqs. (3) and (4).

In contrast to the intrinsic difficulties of the empirical models discussed in the previous section, the extended molecular connectivity model [Eq. (4)] presented in this study lacks these difficulties and is unbiased. No hydrocarbon, halogenated hydrocarbon, or phenol, and heterocyclic and substituted PAH whose soil sorption coefficient has been measured and reported has been omitted from this study. Again, this contrasts with the empirical models (*3,8–13,24*) where only a small subset were considered of the published experimental results on those compounds. In summary, the complete analysis of the molecular connectivity model versus two most commonly used empirical models shows that the molecular connectivity model is clearly superior in accuracy, performance, and range of applicability to empirical models based on *n*-octanol/water partition coefficients or water solubility.

Our molecular connectivity model for predicting soil sorption coefficients was based on nonpolar and nonionic compounds. Thus, it was reasonable to assume that it may not be valid for the highly polar and ionic compounds. To check this assumption our model [Eq. (4)] was tested on the following classes of compounds: anilines, acetanilides, nitrobenzenes, carbamates, substituted benzenes and pyridines, phenylureas, 3-phenyl-1-methylureas, 3-phenyl-1,1-dimethylureas, 3-phenyl-1-cycloalkylureas, alkyl-*N*-phenylcarbamates, triazines, uracils, organic acids, and organic phosphates. The list of compounds sorted by functional groups and their soil sorption coefficients are given by Sabljić (*14*). (In addition to the 143 compounds listed by Sabljić, 16 compounds were used in the initial analysis that could not be simply classified by their functional groups.) The first-order molecular connectivity indices were calculated for all those compounds and their soil sorption coefficients predicted from the molecular connectivity model [Eq. (4)]. We expected the predicted soil sorption coefficients of such a large number of compounds to be distributed randomly around the regression line defined by Eq. (4).

To our surprise, the predicted soil sorption coefficients of all compounds used in this analysis fall below the regression line. Such an unusual result reveals valuable information about the relation between molecular structure and soil sorption properties of organic compounds. First, hydrocarbons must have optimal geometric and electronic features for strong sorption on soils. Second, the introduction of any polar atom and/or substituent will always decrease the soil sorption capability of the resulting compound. Although the second relation was previously known to exist, this is the first systematic and general proof for it and, as shown below, it can be described quantitatively.

The detailed analysis of the calculated soil sorption coefficients of polar compounds shows that the absolute difference between the calculated and observed coefficients depends strongly on the polar functional group and that variations within groups are small. This result is shown in Table 5, where the 143 polar compounds are arranged into 17 classes of compounds in descending order of their soil sorption coefficients. The average error within each functional group shown in Table 5 is very small and is always less than the standard deviation of the molecular connectivity model. Such systematic behavior of the polar organic compounds can be used to indicate the presence or absence of a polar functional group. In addition, the numerical difference between calculated and observed soil sorption coefficients (Table 5, column 3) were used to determine the magnitude of a set of semiempirical variables (the polarity correction factors—P_f) that can be employed to predict accurately the soil sorption coefficients of polar organic compounds. (The only exception to this rule are organic phosphates, which fall into two distinctly different groups. Currently, we are unable to find a structural property that can differentiate between these two groups.) The po-

Table 5. Differences between observed soil sorption coefficients and those predicted from molecular connectivity indices for 143 polar and ionic organic compounds that are used as the polarity correction factors (P_f).

Chemical group	Number of compounds	P_f	Average deviation
Substituted benzenes and pyridines	6	1.00	0.16
Nitrobenzenes	6	1.00	0.24
Anilines	12	1.05	0.20
Organic phosphates (group 1)	9	1.07	0.33
Carbamates	5	1.11	0.11
Phenylureas	15	1.71	0.30
Alkyl-N-phenylcarbamates	10	1.87	0.14
Acetanilides	17	1.87	0.26
3-Phenyl-1-methylureas	5	1.89	0.28
Triazines	11	1.95	0.16
3-Phenyl-1-methyl-1-methoxyureas	4	2.04	0.18
Uracils	3	2.04	0.24
3-Phenyl-1,1-di-methylureas	16	2.23	0.24
Dinitrobenzenes	7	2.36	0.36
Organic acids	8	2.46	0.22
3-Phenyl-1-cycloalkylureas	4	2.57	0.14
Organic phosphates (group 2)	5	3.05	0.28

larity correction factor is calculated for each class of polar compounds by Eq. (5).

$$P_f = \frac{1}{n} \sum \left[\log K_{om}^{eq4} - \log K_{om}^{obs} \right] \quad (5)$$

where n is the number of compounds in particular class. The soil sorption coefficients of polar compounds were recalculated using the polarity correction factors by Eq. (6).

$$\log K_{om}^{adj} = \log K_{om}^{eq4} - P_f \quad (6)$$

The quantitative relationship between recalculated and observed soil sorption coefficients for a total of 215 compounds is described by the regression Eq. (7) and its statistics.

$$\log K_{om}^{obs} = 0.966 \log K_{om}^{adj} + 0.078$$

$$n = 215 \qquad r = 0.969 \qquad s = 0.279 \quad (7)$$
$$F^{1,213} = 3291 \qquad EV = 93.9\%$$

It is now possible, with this very simple model, to accurately predict the soil sorption coefficients for almost 95% of all organic chemicals whose soil sorption coefficients have been measured. No empirical or nonempirical model has ever predicted the soil sorption coefficients to such a high level of accuracy on such a broad selection of structurally diverse compounds.

Mechanism(s) of the Soil Sorption Process

It has long been assumed that soil sorption of nonionic organic compounds could be described as the physical adsorption of solutes on soil organic matter ($1,54-56$). Adsorption can be viewed as a two-dimensional process in which the adsorbed molecules are assumed to be in the plane of the solid's surface. A second model for adsorption is that of a three-dimensional interfacial region bordering on the solid surface where the solute has different thermodynamic properties from those in the bulk phase because of the effect of the solid surface. More recently, two groups ($8,12$) have proposed an alternative concept of soil sorption in which solute partition takes place between the aqueous and the soil organic matter phases. (Partition in the present context is defined as the ratio between the activities of a solute in two bulk phases in equilibrium.) The supporting evidence for the partition process includes a) the linearity of soil-water equilibrium isotherms, b) a very good correlation between the soil sorption coefficients and water solubility of nonionic organic compounds, and c) thermodynamic arguments. It is obvious that this evidence is interpretative, speculative, and may be disputed and/or reinterpreted. This is exactly what happened during the past several years ($17,57$). We will discuss the validity of the above evidence and the feasibility of soil sorption being a partition process.

The first argument put forward in favor of a partition is the linearity of soil-water equilibrium isotherms for nonionic chemicals, even at high, relative concentrations. However, Karickhoff's group (12) measured the soil sorption of 10 small polycyclic aromatic hydrocarbons (including benzene) and chlorinated hydrocarbons and found that "as the sorbate water concentration approaches 60 to 70% of the sorbate aqueous solubility, the isotherms typically bend upward, indicative of increased sorption." Similar nonlinear behavior was found for dibromoethylene, parathion, hexachlorocyclohexanes, napropamide, and terbufos. These results are collected in the study by Mingelgrin and Gerstl (17). In addition, linear adsorption isotherms of parathion, monuron, methyl-parathion, oxydipropionitril, trietazine, hexanol, hexachlorocyclohexanes, and glycine and its peptides on clays was reported ($17,58$). Such adsorbents are very different from the soil organic matter; they do not possess the presumed solvent action of the organic matter fraction that is invoked as the phase where partition from the aqueous solution occurs in soil. Thus, it is fair to conclude that this argument is not generally valid for soil sorption of nonionic chemicals, and it supports both the adsorption and partition processes.

The second argument favoring partition is the very good correlation between the soil sorption coefficients and the water solubility of nonionic organic compounds. We demonstrated in the third section that a very good correlation between the soil sorption coefficients and water solubility of nonionic organic compounds looks more like wishful thinking than reality itself. The truth is that the correlation holds only within a congeneric series of compounds. Thus, serious caution must be exercised not to assign general applicability to them. For example, correlation based on polycyclic aromatic compounds gives reasonable estimates for soil sorption coef-

ficients of small chlorinated hydrocarbons, but tends to overestimate the sorption of highly chlorinated hydrocarbons by as much as an order of magnitude (*16*).

Finally, a weak correlation (r^2 = 0.635) is obtained (*17*) between K_{om} and WS for a large number of nonionic, structurally diverse pesticides. In any case, the lack of good correlation between these two quantities does not prove or disprove the adsorption or partition models. Note here that better correlations are usually found between the soil sorption coefficients and water solubility than between the soil sorption coefficients and *n*-octanol/water partition coefficients. If the soil sorption process is mimicking a partition process, we would expect the opposite to be true.

The third argument put forward favoring partition being the process of soil sorption is the thermodynamic effects. Chiou, Peters, and Freed argued that the observed enthalpies support the partition hypothesis (*8*). This argument is based on an assumption that the adsorption process implies a decrease in entropy and requires a high enthalpy to make adsorption feasible. By the same reasoning, partition may not be as exothermic as adsorption.

In reality, these assumptions are not valid. Adsorption from solutions is usually a competitive process. The entropy can therefore be either positive or negative, depending on the balance between the entropy of the solvent and that of the solute. Consequently, the change in enthalpy of adsorption may have any magnitude and sign. Since the changes in enthalpy and entropy for both partition and adsorption may vary significantly in magnitude and sign, the value of these parameters does not prove or disprove the adsorption or partition models.

Below we examine our results on modeling soil sorption of nonionic organic chemicals with regard to the nature of this process, namely, surface adsorption or solute partition. We proposed (*28*) that the first-order molecular connectivity index be viewed as a quantitative measure of the area occupied by the optimal projection of the molecular skeleton. Since the large majority of the 72 studied nonpolar compounds are "quasi" two dimensional, their index should correlate very well with their surface areas. This prediction was found (*14,28*) to be true for all nonpolar compounds discussed in this study. The molecular surface areas (S_M) of those compounds were calculated by the method recently suggested by Gavezzotti (*59*). The following regression equation is obtained for all 72 compounds:

$$S_M = 24.6 \; ^1\chi + 57.7 \qquad (8)$$

with the correlation coefficient equal to 0.956. This result supports the proposed physical meaning of the first-order molecular connectivity index. Since the two-dimensional representation of the molecular structure is sufficient to explain all quantitative variations in the soil sorption data of nonpolar compounds, the process of soil sorption may be viewed as an attractive interaction between two planes, with the magnitude directly proportional to the main plane area of the solute molecule. Thus, it can be concluded that our results on modeling the soil sorption process favor an adsorption mechanism.

The available results and the previous discussion clearly show that the arguments presented (*8*) in favor of the partition process being the physical basis of soil sorption of nonionic organic molecules are either unjustified or insufficient. It is unfortunate that such a partial study (*8*) on the mechanism of soil sorption was published in the highly regarded scientific journal *Science*, giving it, in our opinion, undeserved respectability. Only a few (*17,57,60*) studies have made a critical assessment of their arguments, while the majority of published articles uncritically accept the partition as a proven, universal physical basis for the soil sorption of nonionic organic molecules (*2,3,9-13,16,24-26,47,61*). At any rate, the available results cannot be used as an unambiguous support of an adsorption mechanism. We speculate that it is most likely that the complexity of the soil and the process of soil sorption are so great that in reality many interactions may control the process, and it is impossible to assign a single mechanism to it. Consequently, the mechanism of soil sorption will be the balance of several possible interactions between the soil and solute molecule, the magnitude of which will be decided by the structural characteristics of the solute molecule and the composition of the soil.

Obviously, much work is needed before clearly identifying specific mechanisms of soil sorption is possible. One possible direction, increasing the fundamental understanding of the mechanisms involved in soil sorption of nonionic chemicals, is to study molecular modeling and computer simulations of this process. The main obstacle to this approach is the lack of an accurate molecular model of soil organic matter. The significant fraction of soil organic matter is humic, thus a solid starting point will be creating an accurate molecular picture of humic acids. The methodology of such an approach should include highly sophisticated computer graphics routines, coupled with molecular mechanics or/and quantum mechanical methods. Of course, the experience on modeling the interactions with or within macromolecules (polypeptides, DNA, RNA, etc.) will be of enormous help.

Closing Remarks

Areas of Application for Molecular Connectivity Model

In this final section we briefly review the areas of application and planned improvements for the molecular connectivity model.

This topological model outperforms the models based on the *n*-octanol/water partition coefficients or water solubility in the accuracy, speed, and range of applicability. In addition, direct correspondence between molecular structure and this particular topological index makes it possible to locate structural features respon-

sible for environmental behavior of organic pollutants in the soil and learn more about the underlying mechanism(s) of soil sorption processes on a molecular level. With the addition of a single semiempirical variable, the presented topological model is also very useful for predicting soil sorption coefficients of polar and even ionic compounds. In fact, we were able to predict accurately the soil sorption coefficients for nearly 95% of all organic chemicals whose soil sorption coefficients have been measured. This result gives us confidence that this topological model will also be accurate in predictions for the compounds whose soil sorption coefficients have not yet been measured. Such an accurate, simple, and fast nonempirical model is almost the ideal predictive tool for ranking potentially hazardous chemicals and for creating priority lists for testing them. The model will enable government agencies, industry, and public groups to make fast and accurate assessments on the environmental fate of proven or potentially hazardous chemicals. The practical need for such a reliable, fast nonempirical model is particularly emphasized by the new U.S. Federal Pesticides Law which will require the review of nearly 600 active pesticide ingredients for their environmental safety within 2 years after the law is passed by Congress.

Finally, the nonempirical nature of this model will enable the manufacturers of pesticides and other classes of organic pollutants to predict accurately the soil sorption capacity and, consequently, the potential environmental hazards of their future products, even before such compounds are synthesized.

The second area of application for molecular connectivity model are exposure assessment methods for groundwater quality (62,63). It should be possible to couple our model for estimating sorption coefficients with the models for calculating concentration profiles and the transport of chemicals in groundwater aquifers. This will result in better assessment methods for predicting exposure concentrations of xenobiotic chemicals in groundwater resources. Predictions of chemical concentration with a high degree of accuracy are essential for retrospective and prospective epidemiological studies, for modification of agricultural practices to reduce the quantity of these chemicals entering groundwater resources, and for the reclamation of contaminated aquifers.

In the future, the molecular topology approach may also be helpful in developing new global ecological models that will give more insight into the expected distribution patterns and the behavior of chemicals in the environment. The molecular topology approach relative simplicity and the possibility to reduce substantially the empirical component in ecological modeling studies make it far more atractive than the correlations between empirical variables, which are of limited usefulness because they lack theoretical basis. In ecological models, different distribution coefficients are combined to describe a type of ecosystem. It would be extremely convenient to have quantitative models describing various environmental distribution coefficients of commercial chemicals based on the same or similar concepts. This will facilitate the development of global, complex environmental models since they can be easily built up as combinations of modules, with each module representing an individual environmental property.

Since the molecular connectivity indices were found to correlate with the soil sorption coefficients (14,27–30,40,41), bioconcentration factors (29,34,40–42,48), biodegradability (50), and acute toxicity (29,35,40, 41,43,45,46,49) of xenobiotic chemicals, it should be feasible from molecular topology alone to evaluate the overall distribution patterns and possible adverse effects of chemicals in ecosystems. This approach will also help us to understand better the underlying mechanisms of how molecular structure influences environmental behavior of organic chemicals.

Future Developments for Molecular Connectivity Model

It is unfortunate that we had to introduce a semiempirical variable into our model and lose the beauty of a purely nonempirical model, trying to extend its use for polar and ionic chemicals. Thus, our future efforts will be focused on finding the structural variable(s) which can explain and quantify the soil sorption behavior of polar and ionic chemicals. Structural analysis of our results on polar and ionic compounds shows that their soil sorption capacity depends strongly on the presence or absence of particular polar functional group(s), and other factors have only minor influence on resulting soil sorption coefficient.

The next developmental stage for the molecular connectivity model is to test its ability to predict the soil sorption coefficients of new, commercial chemicals that are registered at a very high rate. Their main structural characteristics are unusually large size and simultaneous presence of a large number and variety of functional groups. Such a trend will be emphasized even more in the future. Thus, it is important to learn whether or not our model will work correctly for these chemicals.

Our model, as well as all other published models, are developed for relatively small and functionally simple chemicals and may be invalid for large and highly complex molecular structures. Such situations will require either introducing new structural descriptors to account for the unique behavior of the new commercial chemicals, or developing a completely new way of reasoning about these chemicals. The latter solution, if needed, will be an exceptionally difficult task with respect to efforts and time.

Another area of interest in the field of environmental modeling is to develop molecular connectivity models that will predict sorption properties of commercial chemicals on other soil or sediment components (clay, sand, silt, swelling clay, etc.) and the wide variety of surfaces frequently encountered in the subsurface. When developed, these models will be extremely valuable, since the experimental data for such materials are exceptionally scarce. The calculated sorption coef-

ficients will be incorporated into the transport models (*62,63*) to calculate concentration profiles for a wide variety of field conditions. These examples will be also used to identify sources of error that are introduced by the numerical methods used in transport models and the limitations of models used to calculate sorption coefficients from structural characteristics. Consequently, steps will be taken to eliminate or minimize their impact and develop better models.

REFERENCES

1. Hance, R. J. Herbicides and the soil. Chem. Br. 16: 128–131 (1980).
2. Kenaga, E. E., and Goring, C. A. I. Relationship between water solubility, soil sorption, octanol-water partitioning, and concentration of chemicals in biota. In: Aquatic Toxicology (J. G. Eaton, P. R. Parrish, and A. C. Hendricks, Eds.), American Society for Testing and Materials, Philadelphia, PA, 1980, pp. 78–115.
3. Chiou, C. T., Porter, P. E., and Schmedding, D. W. Partition equilibria of nonionic organic compounds between organic matter and water. Environ. Sci. Technol. 17: 227–231 (1983).
4. EPA. Waste Disposal Practices and Their Effects on Groundwater. Report to Congress. U.S. EPA, Washington, DC, 1977.
5. CEQ. Environmental Quality–1980. Eleventh Annual Report of the Council of Environmental Quality, Washington, DC, 1980.
6. EPA. Proposed Groundwater Protection Strategy. U.S. EPA, Office of Drinking Water, Washington, DC, 1980.
7. Connell, D. W. Ecotoxicology—a framework for investigation of hazardous chemicals in the environment. Ambio 16: 47–50 (1987).
8. Chiou, C. T., Peters, L. J., and Freed, V. H. A physical concept of soil-water equilibria for nonionic organic compounds. Science 206: 831–832 (1979).
9. Briggs, G. G. Theoretical and experimental relationships between soil adsorption, octanol-water partition coefficients, water solubilities, bioconcentration factors, and the parachor. J. Agric. Food Chem. 29: 1050–1059 (1981).
10. Schwarzenbach, R. P., and Westall, J. Transport of nonpolar organic compounds from surface water to groundwater. Laboratory sorption studies. Environ. Sci. Technol. 15: 1360–1367 (1981).
11. Schellenberg, K., Leuenberger, Ch., and Schwarzenbach, R. P. Sorption of chlorinated phenols by natural sediments and aquifer materials. Environ. Sci. Technol. 18: 652–657 (1984).
12. Karickhoff, S. W., Brown, D. S., and Scott, T. A. Sorption of hydrophobic pollutants on natural sediments. Water Res. 13: 241–248 (1979).
13. Means, J. C., Wood, S. G., Hassett, J. J., and Banwart, W. L. Sorption of amino- and carboxy-substituted polynuclear aromatic hydrocarbons by sediments and soils. Environ. Sci. Technol. 16: 93–98 (1982).
14. Sabljić, A. On the prediction of soil sorption coefficients of organic pollutants from molecular structures: application of molecular connectivity model. Environ. Sci. Technol. 21: 358–366 (1987).
15. Lyman, W. J. Estimation of physical properties. In: Environmental Exposure from Chemicals, Vol. 1 (W. B. Neely and G. E. Blau, Eds.), CRC Press, Boca Raton FL, 1985, pp. 13–47.
16. Karickhoff, S. W. Organic pollutant sorption in aquatic systems. J. Hydraulic Eng. 110: 707–735 (1984).
17. Mingelgrin, U. and Gerstl, Z. Reevaluation of partitioning as a mechanism of nonionic chemicals adsorption in soils. J. Environ. Qual. 12: 1–11 (1983).
18. Irvin, J. A., and Quickenden, T. I. Linear least squares treatment when there are errors in both *x* and *y*. J. Chem. Educ. 60: 711–712 (1983).
19. Maes, F. W. Problems with linear regression analysis of pharmacological and biochemical data. J. Theor. Biol. 111: 817–819 (1984).
20. York, D. Least-squares fitting of a straight line. Can. J. Phys. 44: 1079–1086 (1966).
21. Hansch, C., and Leo, A. Substituent Constants for Correlation Analysis in Chemistry and Biology. Wiley and Sons, New York, 1979.
22. Rekker, R. F., and de Kort, H. M. The hydrophobic fragmental constant; an extension to a 1000 data point set. Eur. J. Med. Chem. 14: 479–488 (1979).
23. Yalkowsky, S. H., and Valvani, S. C. Solubilities and partitioning: 2. Relationships between aqueous solubilities, partition coefficients, and molecular surface areas of rigid aromatic hydrocarbons. J. Chem. Eng. Data 24: 127–129 (1979).
24. Means, J. C., Wood, S. G., Hassett, J. J., and Banwart, W. L. Sorption of polynuclear aromatic hydrocarbons by sediments and soils. Environ. Sci. Technol. 14: 1524–1528 (1980).
25. Banwart, W. L., Hassett, J. J., Wood, S. G., and Means, J. C. Sorption of nitrogen-heterocyclic compounds by soils and sediments. Soil Sci. 133: 42–47 (1982).
26. Wijayaratne, R. D., and Means, J. C. Sorption of polycyclic aromatic hydrocarbons by natural estuarine colloids. Mar. Environ. Res. 11: 77–89 (1984).
27. Sabljić, A., and Protić, M. Relationship between molecular connectivity indices and soil sorption coefficients of polycyclic aromatic hydrocarbons. Bull. Environ. Contam. Toxicol. 28: 162–165 (1982).
28. Sabljić, A. Predictions of the nature and strength of soil sorption of organic pollutants by molecular topology. J. Agric. Food Chem. 32: 243–246 (1984).
29. Sabljić, A. Nonempirical modeling of environmental distribution and toxicity of major organic pollutants. In: QSAR in Environmental Toxicology—II (K. L. E. Kaiser, Ed.), D. Reidel Publishing Co., Dordrecht, Holland, 1987, pp. 309–332.
30. Trinajstić, N. Chemical Graph Theory, Vols. I and II. CRC Press, Boca Ration, FL, 1983.
31. Sabljić, A., and Trinajstić, N. Quantitative structure-activity relationships: the role of topological indices. Acta Pharm. Jugosl. 31: 189–214 (1981). Program GRAPH III is now fully operational and it is available for distribution.
32. Randić, M. On characterization of molecular branching. J. Am. Chem. Soc. 97: 6609–6615 (1975).
33. Kier, L. B., and Hall, L. H. Molecular Connectivity in Chemistry and Drug Research. Academic Press, New York, 1976.
34. Sabljić, A., and Protić, M. Molecular connectivity: a novel method for prediction of bioconcentration factor of hazardous chemicals. Chem.-Biol. Interact. 42: 301–310 (1982).
35. Sabljić, A. Quantitative structure-toxicity relationship of chlorinated compounds: a molecular connectivity investigation. Bull. Environ. Contam. Toxicol. 30: 80–83 (1983).
36. Sabljić, A. Calculation of retention indices by molecular topology: chlorinated alkanes. J. Chromatogr. 314: 1–12 (1984).
37. Sabljić, A. Calculation of retention indices by molecular topology: chlorinated benzenes. J. Chromatogr. 319: 1–8 (1985).
38. Sabljić, A., and Protić-Sabljić, M. Quantitative structure-activity study on the mechanism of inhibition of microsomal p-hydroxylation of aniline by alcohols. Role of steric factors. Mol. Pharmacol. 23: 213–218 (1983).
39. Šoškić, M., and Sabljić, A. Inhibition of hill reaction by 3-alcoxyuracil derivatives: QSAR study with topological indices. Croat. Chim. Acta 60: 755–764 (1987).
40. Koch, R. Molecular connectivity index for assessing ecotoxicological behavior of organic compounds. Toxicol. Environ. Chem. 6: 87–96 (1983).
41. Govers, H., Ruepert, C., and Aiking, H. Quantitative structure-activity relationships for polycyclic aromatic hydrocarbons: correlation between molecular connectivity, physico-chemical properties, bioconcentration and toxicity in *Daphnia pulex*. Chemosphere 13: 227–236 (1984).
42. Sabljić, A. The prediction of fish bioconcentration factors of organic pollutants from the molecular connectivity model. Z. Gesamte Hyg. Grenzgeb. 33: 493–496 (1987).
43. Mekenyan, O., Bonchev, D., Sabljić, A., and Trinajstić, N. Applications of topological indices to QSAR. The use of the Balaban index and the electropy index for correlations with toxicity of ethers on mice. Acta Pharm. Jugosl. 37: 75–86 (1987).
44. Randić, M., Jeričević, Ž., Sabljić, A., and Trinajstić, N. On the

molecular connectivity and π-electronic energy in polycyclic hydrocarbons. Acta Physiol. Pol. A74: 317–330 (1988).

45. Basak, S. C., Gieschen D. P., and Magnuson, V. R. A quantitative correlation of the LC50 values of ethers in *Pimephales promelas* using physicochemical and topological parameters. Environ. Toxicol. Chem. 3: 191–199 (1984).

46. Craig, P. N., and Enslein, K. Structure-activity in hazard assessment. In: Hazard Assessment of Chemicals, Vol 1, Current Developments (J. Saxena and F. Fisher, Eds.), Academic Press, New York, 1981, pp. 389–420.

47. Koch, R. A theoretical-methodological approach towards the evaluation of environmental pollutants. Toxicol. Environ. Chem. 7: 331–346 (1984).

48. Vighi, M., and Calamari, D. QSAR for organotin compounds on *Daphnia magna*. Chemosphere 14: 1925–1932 (1985).

49. Yoshioa, Y., Mizuno, T., Ose Y., and Sato T. The estimation for toxicity of chemicals on fish by physico-chemical properties. Chemosphere 15: 195–203 (1986).

50. Boethling, R. S. Application of molecular topology to quantitative structure-biodegradation relationships. Environ. Toxicol. Chem. 5: 797–806 (1986).

51. Newsted, J. L., and Giesy, J. P. Predictive models for photoinduced acute toxicity of polycyclic aromatic hydrocarbons to *Daphnia magna*. Environ. Toxicol. Chem. 6: 445–461 (1987).

52. Balaban, A. T., Motoc, I., Bonchev, D., and Mekenyan, O. Topological indices for structure-activity correlations. Top. Curr. Chem. 114: 21–55 (1983).

53. Seybold, P. G., May, M., and Bagal, U. A. Molecular structure-property relationships. J. Chem. Educ. 64: 575–581 (1987).

54. Furmidge, C. G. L., and Osgerry, J. M. Persistence of herbicides in soil. J. Sci. Food Agric. 18: 269–273 (1967).

55. Bailey, G. W., and White, J. L. Factors influencing the adsorption, desorption, and movement of pesticides in soil. Residue Res. 32: 29–92 (1970).

56. Carringer, R. D., Weber, J. B., and Monaco, T. J. Adsorption-desorption of selected pesticides by organic matter and montmorillonite. J. Agric. Food Chem. 23: 568–572 (1975).

57. Kyle, B. G. Soil-water equilibria for nonionic organic compounds. Science 213: 683–684 (1981).

58. Greenland, D. J., Laby, R. H., and Quirk, J. P. Adsorption of glycine and its di-, tri-, and tetra-peptides by montmorillonite. Faraday Soc. Trans. 58: 829–841 (1962).

59. Gavezzotti, A. Molecular free surface: a novel method of calculation and its uses in conformational studies and in organic crystal chemistry. J. Am. Chem. Soc. 107: 962–967 (1985).

60. MacIntyre, W. G., and Craig, L. S. Comment on "Partition equilibria of nonionic organic compounds between soil organic matter and water." Environ. Sci. Technol. 18: 295–297 (1984).

61. Dzombak, D. A., and Luthy, R. G. Estimating adsorption of polycyclic aromatic hydrocarbons on soils. Soil Sci. 137: 292–308 (1984).

62. Lindstrom, F. T., and Piver, W. T. Vertical transport and fate of low-water-solubility chemicals in unsaturated soils. J. Hydrol. 82: 93–141 (1985).

63. Lindstrom, F. T., and Piver, W. T. Vertical-horizontal transport and fate of low water solubility chemicals in unsaturated soils. J. Hydrol. 86: 93–131 (1986).